CLAIMS

1. A method for discovering knowledge from text documents, the method comprising the steps of:

extracting from text documents semi-structured meta-data, wherein the semi-structured meta-data includes a plurality of entities and a plurality of relations between the entities;

identifying from the semi-structured meta-data a plurality of key entities and a corresponding plurality of key relations;

deriving from a domain knowledge base a plurality of attributes relating to each of the plurality of entities relating to one of the plurality of key entities for forming a plurality of pairs of key entity and a plurality of attributes related thereto;

formulating a plurality of patterns, each of the plurality of patterns relating to one of the plurality of pairs of key entity and a plurality of attributes related thereto;

analyzing the plurality of patterns using an associative discoverer; and

interpreting the output of the associative discoverer for discovering knowledge.

2. The method as in claim 1, wherein the step of extracting from text documents comprises the step of extracting text content from documents containing at least one type of text, image, audio, and video information.

3. The method as in claim 1, wherein the step of identifying the plurality of key entities comprises the step of selecting the plurality of key entities according to at least one of frequency of appearance of the plurality of key entities in the semi-structured meta-data and obtaining user specification.

4. The method as in claim 1, wherein the step of identifying the plurality of key relations comprises the step of selecting the plurality of key relations according to at least one of frequency of appearance of the plurality of key relations in the semi-structured meta-data and obtaining user specification.

5. The method as in claim 1, wherein the step of deriving from the domain knowledge base comprises the step of deriving from a domain knowledge base relating to at least one of taxonomy, a concept hierarchy network, ontology, a thesaurus, a relational database, and an object-oriented database.

6. The method as in claim 1, wherein the step of deriving the plurality of attribute comprises the step of deriving a set of attributes or lower level entities characterizing the plurality of entities relating to the plurality of key entities.

7. The method as in claim 1, wherein step of the formulating the plurality of patterns comprises the step of formulating concatenated vector representations of the plurality of attributes and the plurality of key entities relating to the corresponding plurality of key relations.

8. The method as in claim 1, wherein the step of analyzing the plurality of patterns using the associative discoverer comprises the step of analyzing the plurality of patterns using at least one of a neural network, a statistical system, and a symbolic machine learning system.

9. The method as in claim 8, wherein the step of analyzing the plurality of patterns comprises the step of analyzing the plurality of patterns using an Adaptive Resonance Associative Map.

10. The method as in claim 1, wherein the step of interpreting the output of the associative discoverer for discovering knowledge comprises the step of discovering the relations between the plurality of attributes and the plurality of key entities.

11. The method as in claim 1, further comprising the step of using a user interface for displaying the semi-structured meta-data, the plurality of key entities, the plurality of key relations, the plurality of attributes, and the knowledge discovered.

12. The method as in claim 1, further comprising the step of using a user interface for obtaining user instruction for the plurality of key entities and the plurality of key relations.

13. A computer program product comprising a computer usable medium having computer readable program code means embodied in the medium for discovering knowledge from text documents, the computer program product comprising:

computer readable program code means for extracting from text documents semi-structured meta-data, wherein the semi-structured meta-data includes a plurality of entities and a plurality of relations between the entities;

computer readable program code means for identifying from the semi-structured meta-data a plurality of key entities and a corresponding plurality of key relations;

computer readable program code means for deriving from a domain knowledge base a plurality of attributes relating to each of the plurality of entities relating to one of the plurality of key entities for forming a plurality of pairs of key entity and a plurality of attributes related thereto;

computer readable program code means for formulating a plurality of patterns, each of the plurality of patterns relating to one of the plurality of pairs of key entity and a plurality of attributes related thereto;

computer readable program code means for analyzing the plurality of patterns using an associative discoverer; and

computer readable program code means for interpreting the output of the associative discoverer for discovering knowledge.

14. The computer program product as in claim 13, wherein the computer readable program code means for extracting from text documents comprises computer readable program code means for extracting text content from documents containing at least one of text, image, audio, and video information.

15. The computer program product as in claim 13, wherein the computer readable program code means for identifying the plurality of key entities comprises computer

readable program code means for selecting the plurality of key entities according to at least one of frequency of appearance of the plurality of key entities in the semi-structured meta-data and obtaining user specification.

16. The computer program product as in claim 13, wherein the computer readable program code means for identifying the plurality of key relations comprises computer readable program code means for selecting the plurality of key relations according to at least one of frequency of appearance of the plurality of key relations in the semi-structured meta-data and obtaining user specification.

17. The computer program product as in claim 13, wherein the computer readable program code means for deriving from the domain knowledge base comprises computer readable program code means for deriving from a domain knowledge base relating to at least one of taxonomy, a concept hierarchy network, ontology, a thesaurus, a relational database, and an object-oriented database.

18. The computer program product as in claim 13, wherein the computer readable program code means for deriving the plurality of attributes comprises computer readable program code means for deriving a set of attributes or lower level entities characterizing the plurality of entities relating to the plurality of key entities.

19. The computer program product as in claim 13, wherein the computer readable program code means for formulating the plurality of patterns comprises computer readable program code means for formulating concatenated vector representations of the plurality of attributes and the plurality of key entities relating to the corresponding plurality of key relations.

20. The computer program product as in claim 13, wherein the computer readable program code means for analyzing the plurality of patterns using the associative discoverer comprises computer readable program code means for analyzing the plurality of patterns using at least one of a neural network, a statistical system, and a symbolic machine learning system.

21. The computer program product as in claim 20, wherein the computer readable program code means for analyzing the plurality of patterns comprises computer readable program code means for analyzing the plurality of patterns using an Adaptive Resonance Associative Map.

22. The computer program product as in claim 13, wherein the computer readable program code means for interpreting the output of the associative discoverer for discovering knowledge comprises computer readable program code means for discovering the relations between the plurality of attributes and the plurality of key entities.

23. The computer program product as in claim 13, further comprising computer readable program code means for using a user interface for displaying the semi-structured meta-data, the plurality of key entities, the plurality of key relations, the plurality of attributes, and the knowledge discovered.

24. The computer program product as in claim 13, further comprising computer readable program code means for using a user interface for obtaining user instruction for the plurality of key entities and the plurality of key relations.

25. A system for knowledge discovery from free-text documents, comprising:

means for extracting semi-structured meta-data from the free-text documents;

means for identifying key entities and key relations from the semi-structured meta-data;

a knowledge base that defines the attributes of entities;

means for formulating patterns based on the key entities and the attributes of entities related to the key entities; and

means for analyzing the patterns for knowledge.

26. The system according to claim 25 wherein the semi-structured meta-data comprises definition of entities and relations among the entities.

27. The system according to claim 25 wherein the semi-structured meta-data is stored in a permanent or temporary storage.

28. The system according to claim 25 wherein the free-text documents comprise text, image, audio, video, or any combination thereof.

29. The system according to claim 25 wherein the means for identifying key entities selects entities according to at least one of the key entities' frequency of appearance in the semi-structured meta-data and user's specification.

30. The system according to claim 25 wherein the means for identifying key relations selects relations according to at least one of the key relations' frequency of appearance in the semi-structured meta-data and user's specification.

31. The system according to claim 25 wherein the knowledge base comprises a taxonomy, a concept hierarchy network, an ontology, a thesaurus, a relational database, an object-oriented database, or any combination thereof.

32. The system according to claim 25 wherein the attributes of entities comprise a set of attributes or lower level entities characterizing the entities.

33. The system according to claim 25 wherein the training examples comprises concatenated vectors of the key entities, and the attributes of entities related to the key entities with a key relation.

34. The system according to claim 25 wherein the pattern analyzer comprises a neural network, a statistical system, a symbolic machine learning system, or any combination thereof.

35. The system according to claim 25 wherein the pattern analyzer comprises an Adaptive Resonance Associative Map.

36. The system according to claim 25 wherein the knowledge comprises hidden key relations between the attributes of the entities and the key entities.

37. The system according to claim 25 wherein the knowledge discovery system further comprises a user interface for displaying the semi-structured meta-data, the key entities, the key relations, the attributes, and the knowledge discovered.

38. The system according to claim 25 wherein the knowledge discovery system further comprises a user interface for obtaining user's instruction for the key entities and the key relations.